



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE ANALYSIS OF A CENSUS

BY G. H. KNIBBS, *Federal Statistician of Australia*

1. *General*.—At the present time the results of a census are usually published in the form of a series of tables, with some measure of analysis and interpretation in the form of comment. To a limited extent also, derivative results such as life-tables are sometimes given. But every competent statistician realizes that this is very far from furnishing any adequate idea of the real significance of the data, so labouriously acquired, and necessarily at considerable public expense. Hence the question naturally arises: “What ought to be the form of presentation of the results of future censuses?” This is the question I now propose to discuss.

Precedents already exist for doing much more than what is above suggested. The five professional papers published, for example, in connection with the United States Coast and Geodetic Survey are instances of the kind of thing which might well be done in connection with a census. The work itself in this survey consists of the measurement of base-lines, and of the angles of a network of triangles covering the whole area surveyed, together with the associated astronomical, chronometric and metrological work. This involves researches into the geometry of the earth's figure, into physics and geophysics, into practical astronomy, into the reduction of observations, into optical questions, into the measurement of the force of gravity, and so on. Indeed, one might quite properly say that the mathematical and physical sciences are all called into requisition for the purpose of the survey, and the researches cover a correspondingly wide range.

These researches are published (as already said) in professional papers, constituting not merely the appendix to the formal report, but often its most valuable element. Anyone acquainted with them realizes their value and the wisdom and public utility of the practice.

One might well say that a survey of the human beings occupying any territory is of at least as high an order of importance as the survey of the area which they occupy. Indeed the necessity of such a survey will emerge as soon as anyone attends to the facts; for indeed it must be self-evident that anything short of such a survey will not meet the inherent needs of the case. “The proper study of mankind is Man”; not merely in himself and in relation to his fellows, but also in relation to facts relevant to his well-being on the earth.

In this connection moreover it may be said that our ordinary statistical conceptions are not adequate. We think readily of "averages" but not of "distributions" about average values. The Belgian Astronomer-Mathematician Quetelet laid a foundation,—now quite a considerable number of years ago. But his conceptions, valuable as they were, are only a beginning. The notion of an average has to be supplemented by that of the characteristic of a distribution about the average or mean. And this desideratum has given birth to a new science, viz., the general theory of statistics and statistical results. This theory involves questions of probability, of the mode in which group-numbers arise, of the way in which individually observed facts are grouped (or dispersed) about their mean value, the dispersion or "scatter" itself often being a matter of high importance.

This is not the only element which needs development. We may have adequate grounds for believing that two series of facts are mutually related, and we may require a numerical expression for the degree of relation (correlation). By means of an appropriate analysis this can be obtained: thus we have a theory of correlation, which has wide applications in the Mendelian theory of inheritance.

Or yet again we may have data which, properly examined, enable us to see the law running through a series of phenomena, for example, the frequency of marriage or birth in the case of persons of different pairs of ages, and so on. When we can subsume the whole of the phenomena in question under some definite conception, expressible either by means of an algebraic equation, or by means of a graph, we powerfully advance our knowledge.

2. *Analysis and What It Involves.*—What has been said sufficiently indicates the general point of view. Extensive researches have already been made in various fields. The well-known "*Biometrika*" contains numerous instances both of methodological technique and of analysis. Various papers by Bachelier, Borthkiewicz, Charlier, Davenport, Edgeworth, Lexis, Livi, Pearson, Tschuprow, Yule, Westergaard, Wicksell, and many others, have laid the foundations of a method of great analytical power and importance. And though a comprehensive work on the whole range of statistical methodology is still a desideratum, it would be possible now to construct it. Its creation, however, involves both mathematical knowledge and practical experience in the statistical field, and unfortunately there are but few men who possess both the experience and the necessary leisure. Is not this a work for governments, or worthy of a private benefaction?

The Report on the Australian Census of 1911 contained, in Vol. I, an Appendix entitled "The Mathematical Theory of Population, of its

Character and Fluctuations, and of the Factors which Influence Them." This work (by the author of the present paper)—appreciatively referred to, among others, by American authorities—aimed on the one hand at supplying some necessary elements of mathematical technique, for analyzing various aspects of vital phenomena, and on the other at interpreting material made available by the first Australian Census which had been carried out on uniform lines and by a central authority. It embraced two elements, one technical or methodological, the other interpretative. An examination of it will reveal how current statistical data, supplemented by data rendered available by means of a census, can be made to furnish answers to important questions concerning the increase and character of a population.

It is easy to see that the questionnaire of a census, and the statistical material collected during intercensal periods, must be properly correlated in order to obtain the maximum advantage therefrom. Thus a satisfactory analysis may be said to involve three elements, viz.:

- (i) Mathematical technique and methodology generally;
- (ii) Appropriate correlations of data; and
- (iii) Insight as to the significance of results obtained or obtainable.

The last is the only one to which we need now refer, and it will suffice, for the purpose of illustration, to mention but few elements.

The most fundamental thing about population is its rate of increase. But this depends largely, not merely upon its constitution with respect to sex and age, but upon its food-supply, its economic condition generally, and its standard of living. The last is a powerful factor in altering the rate of possible increase.

We are concerned, however, not only with mere numerical or quantitative results, but also with qualitative; and thus anthropometric measurements which can throw light upon the physique or upon the intellectual ability of a people are of high value. And they are very far-reaching in their consequence. The expansion of races, the difficulties of sufficiently increasing food-supply, the qualitative changes accompanying increasing density of population, the desirability or otherwise of miscegenation; these are subjects in regard to which inquiries are of momentous importance.

On the side of social-economics also, the issues are of no less significance. How national or private wealth is best measured, what are the characteristics of its distribution, how far, if at all, it is desirable that its accumulation should be restricted, how far national services should be nationally controlled or provided for by private effort, how far the levelling up as between nation and nation should be promoted,—these are questions calling for the highest order of statis-

tical analysis, and are questions which may not be postponed without danger.

3. *Mathematics and Its Relation to Practical Results.*—There is a vulgar impression that a statistician is a mere tabulator of data, and that, when tabulations have been effected, the work of the statistician is at an end. As a matter of fact, however, his real work only then commences. And moreover even the tabulation of group-results needs to be directed by one who fully understands the significance of grouping, and what is essential to furnish accurate results in the case of critical facts, such, for example, as the age which corresponds to the highest frequency of marriage, birth, etc., the age of lowest mortality, and so on. But ignoring the last mentioned fact, we may note that crude statistical results are attained with considerable error, and one of the most important functions of the statistician is to eliminate this error as far as it is possible to do so. Thus, an acquaintance with the mathematical theory of error is an essential part of his professional equipment. This in its turn involves a knowledge of the characteristics of chance distributions, for these in many cases furnish a guide for the proper reduction of crude observations so as to furnish the best result obtainable from the imperfect data which come to hand. Again, to obtain the best results one may have to decide whether the crude data should be regarded as a chance-distribution at all, or whether the appropriate conception is that they are functionally related to the values of the variable age, say, in some other way. And this is by no means a simple question, it is one often involving a nice discrimination between points of view.

Again, we may have to ask ourselves exactly what it is that we are seeking to obtain? Is it—for example—that limit which would be attained with an infinitely large population (a sort of ideal result); or is it that which most simply represents a very limited one? And behind these questions lies the whole theory of the empirical representation of numerical facts. We substitute for the irregular results obtained regular ones which conform—if possible—to some readily expressible law. This subsumption of statistical facts under clear-cut conceptions is in conformity with the general method of science, and involves an appropriate command of mathematical theory.

We may take an illustration to make clear what is meant. Let us suppose that the age of most frequent nuptial child-birth is to be ascertained. In a small population the result may be prejudiced by irregularities in the data. Hence we first “smooth” them, getting rid of any zig-zag appearance of the results. Then, only, emerges the ideal curve to which the facts *tend to conform*. This is the essence of the

matter. But to "smooth" results we must know the significance of first, second, and higher differences, or in other words the essential character of the curve which we regard as appropriately representing in their generality the facts examined. We are then, and only then, in a position to indicate with any exactitude the age in question.

It is obvious that life or mortality tables must conform to the requirements indicated in respect of smoothness, and so on. It is, moreover, easy to see that any general scheme of insurance must be based on appropriately *adjusted* data, and results derived therefrom. We need not further labour the point, viz., that mathematical science renders eminently practical service; it is really an essential in the proper analysis, representation, and indeed interpretation also, of statistical results.

4. *Indications of the Range of Statistical Analysis.*—In the first number of "Metron," published at Rome, will be found an article on the "Theory of Large Population Aggregates"* showing the significance of the mere theory of increase of population, and that the increase ultimately becomes zero, viz., when, for any given condition of things, the limiting density of population has been reached. At the rate of population-increase which characterized the last century this would arrive at no very remote date, and in any case the world's future food-supply will soon become a question of the first order of importance. This fact alone will suffice to emphasize the need for the analysis of statistics, and the necessity for entrusting it to experts of appropriate training and powers.

An indication of statistical analyses already carried out may reveal more clearly the real range and importance of the subject.† Starting with sex-ratio, we need to know what is the norm or average at each age for the whole western world, the ratio at birth and generally, for still and live births both nuptial and ex-nuptial, for the first-born, and so on. These facts are of sociologic importance and have a bearing upon national eugenics. Next the birth-rate itself and its relation to the marriage rate require to be ascertained, and where a mixture of races exists, and unrestricted miscegenation is opposed to popular feeling, the relative birth-rate of races within the population becomes itself a matter of far-reaching significance. From the standpoint of national development and from that of national eugenics, the effective birth-rate, after the infantile period—with its special dangers—has passed, is essential. The children saved from infantile disaster are what make up the effective population; in other words, mere birth-rate is itself of but little moment: it is the residual or effective birth-

* By the author of this paper.

† In the Australian Census of 1911, by the author of this paper.

rate with which we are concerned. Thus, Australia and New Zealand, though they have not large birth-rates, are good examples because their residual rates are high: in short, as nations they do not bring children recklessly into existence, but when children do come into being there is an endeavour to save them.

The drift of conjugal frequency and its distribution according to age, as affecting the birth-rate, as well as the stability of society, the inimical effects, and frequency of divorce, constitute a branch of social statistics which will not be neglected by students of sociology, or governments interested in national well-being.

The relative (or absolute) numbers of husbands of a given age, marrying or living with wives of a given age, may be defined by what the writer has called a *protogamic* or a *gamic* surface. The contours of this surface disclose the full conjugal particulars according to pairs of ages at a glance.

For calculating the number of marriages likely to take place, we need a theory of what may be called the *conjugal potential* of a community, and of its variation with changes in their economic condition. The change of probability with differences in the numbers living at each age needs therefore to be considered. Despite the fact that world statistics at present do not afford sufficient data to determine very exactly the form of the function by means of which the influence of variations in the numbers of unmarried persons can be eliminated, a function of high probability has been developed and tabulated, so that the elimination can be effected and the number ascertained.*

In order to represent the birth-rates of two populations in such a way as to eliminate differences in their age-constitution, we must have recourse to *natality indexes*, constituted similarly to mortality indexes.

To ascertain the age representing the beginning and end of fertility we need very extended observations. For Australia it appears to be about as follows:

PROBABILITY OF NUPTIAL MATERNITY

Age.....	11	12	13	14	15	16	...	49	50
	0	0	.5?	.2?	.22	.410088	.0033
Age.....	51	52	53	54	55	56	57	58	
	.0012	.0006	.0005	.0004	.0025	.0015	.0005	.00000	

Obviously, however, these numbers are not those of the ideal curve representing the fact as it would be were it not influenced in some way by limitations of the available data, by social tradition, etc.

* By the writer of this paper: see Appendix I, pp. 214-223.

To express the relative frequency of first births after different durations of marriage and for different ages of women we need also a surface, so that the coördinates are: x =ages of the women; y =the duration of marriage; and z =the frequency. This surface may be called the *nuptial protogenesic surface*, or the *surface of nuptial primiparity*. In the graph of this, the contour for zero is the *agenesic boundary*, and the region outside of this is the *agenesic region*, i. e. it represents the series of ages and durations of marriage which are characterized by sterility. The average interval between marriage and birth for women of a given age may be called the *protogenesic index* for the age in question, and for women of all ages, the *general protogenesic index*.

The insight which this index gives in regard to the Malthusian modification of the physiological condition is of great and obvious sociological importance. It may be mentioned that for reasons which it is not proposed to detail here, it is necessary also to compute what has been called "*the protogenesic quadratic index*."* It may be mentioned that a correction to the calculated protogenesic index is necessary when the population is not constant.† Also the age of greatest fertility is not immediately given by the data, but requires correction for preceding cases of maternity. It is thus for Australia changed from: Age 18.8 and probability 0.483, to Age 19.0 and probability 0.517.

The preceding will afford some general idea of the nature of the analyses of some of the more simple elements which demand consideration. We shall now indicate some of the more complex cases.

5. *More Complex Elements Subject to Statistical Analysis*.—In dealing with statistics of fertility we are confronted with difficulties in regard to data. It has been shown elsewhere‡ that in statistics involving age and duration, where there are partially specified and wholly unspecified cases, these must be distributed according to the laws of probability, and that *unspecified cases follow a regular law*. When this has been done, then and only then can a satisfactory analysis be proceeded with. This being done for Australia it was found that the number of children at a confinement is a definite function of age, and that both for nuptial and ex-nuptial cases, the relative increase with age x up to about age thirty-six is sensibly given by the equation $B = 1 + 0.00065 (x - 12)$ in which B represents the increase in the probability of a multiple birth; after that age the probability decreases again. Similarly the probability of triplets also increases with age, and is about 50 or 60 per cent greater in ex-nuptial than in nuptial

* By the author of this paper; see Appendix referred to pp. 272-274.

† *Ibid.*, pp. 274-275.

‡ *Ibid.*, p. 302.

cases. The probability of twins occurring twice is nearly identical with the occurrence of triplets. The phenomena of ex-nuptial fertility are not very different from those of nuptial except that the maximum occurs about three years later, and tends to be *relatively* greater than the proportion indicated by the maximum from twenty-five years of age onwards.

When the duration of marriage and the age of the mother at the birth are both taken into account, say the ages being the abscissae (x) and the durations of marriage the ordinates (y), the relative frequency being then represented by verticals (z), the surface so defined may be called the *genesic distribution* at maternity. Or if the ages at marriage of the mothers be substituted for the ages at the birth, the distribution may be called the *gamogenesic distribution*. Thus we may adopt either or both for practical purposes, the gamogenesic tables denoting the proportion of wives who, having married at a given age (x), have during a duration of marriage of y years, borne k children.

In the preceding results the age of the husband, etc., is ignored: when taken into account the results become much more complex. The graph of birth-rates, based upon the ages of husbands as abscissae and of wives as ordinates, give a surface which I have called *digenesic*.* The contours defining equal birth-rates are appropriately called *diisogenic contours* and any series of ages giving a particular birth-rate may be called the *diisogenic ages* for that rate. Such surfaces reveal the rapidity with which the effective virility of the husband diminishes with age, and also the effective impairment of the wife, and reveal remarkable differences between different peoples.† Similarly we may construct surfaces for multiple diisogeny.

We shall consider but one other case of complex statistics, viz., that relating to mortality. It has been shown‡ that prediction tables for the frequency of death within a given duration of time, say one year, for persons of a given age are true only for a very limited time indeed. There has recently been a remarkable increase in the probable duration of life for nearly all ages, so great indeed that by the time a "life-table" has been constructed, it is already "out of date." The probable result for a given year can be obtained by extrapolating past results. The curves are sensibly exponential, concave upwards, the probability of mortality, as above said, decreasing with time. The general results show that the diminution in mortality is more marked in the young than in the old, and that it is not identical for males and females. The factors in operation are probably the following:

* See Appendix referred to, p. 350.

† *Ibid.*, pp. 360-361.

‡ By the author of this paper, *Ibid.*, pp. 280-281.

(i) Evolution of the protective reaction between the organism and its environment.

(ii) Changes in the quantity, quality, and appropriateness of the food supply.

(iii) Changes due to the reaction of the organism to economic conditions in respect of its nutritional and neural apparatus.

(iv) Changes in personal and general hygiene, in therapeutics and surgery, in prevailing traditions which affect the vitality of an organism, etc.

I have shown that the facts of the variation of mortality with age suggest the following theory of its nature:

In virtue of its plasticity an organism is both vulnerable and recuperable. At birth the change of environment is inimical, the initial difficulty of adaption constituting the *natal "force of mortality."* This falls off very rapidly indeed, and constitutes what may be called the *natal adaptative effort*. The initial adaption being established in the course of the first few days, there follows what may be called the *infantile adaptative effort* covering, roughly, the first twelve months of life. The inimical forces are now greatly lessened, this period being characterised by the rapidly diminishing *infantile force of mortality*. The growing incapacity for adaptation gives rise to what may be called the *senile force of mortality*, which probably operates from the first. The pressure put upon the organism by social life, by the appearance of puberty, by the assertiveness of the reproductive forces, etc., constitute an inimical force, which vanishes with the termination of the reproductive period, and may be called the *genesic force of mortality*. The exhaustion and general dangers of reproduction also constitute for the female sex an inimical force which may be called the *gestate force of mortality*. There are also dangers of a casual nature, which may in their totality be called the *adventitious force of mortality*. General mortality is the sum of all these.

It is not proposed here to discuss the theory of an actuarial population, or the limitations of the Gompertz-Makeham-Laudi-and-Lazarus theory of mortality. It will suffice to observe that the whole position calls for revision by a competent mathematician.

Enough has been said to indicate that an ordinary population census can be made to tell a far more interesting, instructive, and valuable story than it does at present. It would be easy to multiply examples. To show for instance how the mortality-frequency according to age and sex for various diseases can be obtained, or, carrying the research into the fields of wealth and income, to show how a *plutoprosodic surface* can be made to exhibit the frequency with which particular

incomes are associated with particular amounts of wealth, and so on.

6. *Special Reasons for Analysis at the Present Time.*—Apart from the general importance of statistical analysis in connection with a census, there are quite special reasons why this should be regarded as of more than ordinary moment at the present time. To these reasons we shall now refer.

The widespread movement towards democracy, the assertion of the claims of labour, the challenge of the rights of merely inherited wealth, the financial consequences of the great war, the question of the issues between production for use and production for profit, the anonymity and personal irresponsibility of capital and its reactions on nationalism, the social obligations of all members of any community to the community itself, the rate at which Nature's accumulated resources are being exploited, the possibilities of development of the world's food supply, the general raising of the standards of living, the international advance in eugenics, and similar questions, all indicate the urgency of the task before all great nations. None can afford to neglect to take stock of its affairs, to study its social and economic drift, to forecast its rate of exploitation of its recognised or available resources, to consider how it is to meet the various contingencies of national life, etc. Attention to these matters is indeed a national stock-taking.

Collisions between nations may perhaps be inevitable, but if they can in any measure be avoided, the dictates of reason indicate that this avoidance must be founded upon an intelligent appreciation of the whole of the circumstances of the case, including its relations with other nations. If the economic folly of collision and consequent destruction in war is to be replaced by intelligent and amicable readjustments of national interests, the way lies in a better knowledge of the comparative positions of the affairs of each nation. That is the only "way out."

We are denizens of a world of limitations. Collision of interests may perhaps be ultimately unavoidable, and is certainly inevitable so long as egoism is universal. War and death-struggle may perhaps be involved by the very "nature of things." Or it may not. Be that as it may, one cannot resist the conviction that if one hundredth part of what is spent in war were applied to a study of the problem of solving the difficulties between nation and nation, we should avoid a vast amount of trouble, and approach a partial, if not a complete, solution of international and internecine difficulties.

The proper analysis of a census is a first step towards a really systematic and incisive study of national and world affairs. Without such an

analysis, conjecture, rather than knowledge, governs our thought and our action, and may lead us far astray. The powers at our disposal are becoming continually more terrible, and the "way of force" increasingly ruthless. One can assign no limit to the consequences of appeal thereto, and it may be taken for granted that the next struggle will be of a far more deadly character than that just ended. This surely must give us pause, and should lead all nations to appraise the real state of society, the meaning of individual, of communal and of national life.

There is, awakening in all thoughtful men, a feeling that we shall do well to ask ourselves the great question—"quo vadis". Whither is our civilization trending? What is to be expected if the universal appeal is to the power of inflicting injury?

One may well ask: "Are we in the throes of birth of a new era, and if so what will it witness?" Will it see the world marshalled in dread array in a ghastly appeal to the shrieking Bellona to call forth again the dogs of war, or will it witness a great endeavour to recognize that in that way lies irretrievable disaster and perhaps complete ruin. For be it remembered that we can command, as never before, the vast resources of our common mother Nature? It is not impossible that we may achieve even untold success in new instruments of destruction.

To these questions we may perhaps answer: "Destiny lies on the knees of the gods." Be it so, but we can survey its sway, perhaps can take an intelligent interest, not merely as fatalistic onlookers, but as intelligent coöperators. May it not be our privilege, as statisticians, to point the way in which the maximum safety lies over the uncharted seas of the future? Or if that be the function of the statesman, may we not furnish such a conspectus of international affairs, that the trend of Destiny shall be obvious, and the chart of the World Experience such that we can discern something of the Unknown into which we are continually passing?

I cannot believe that the leaders of intelligent nations will, in future, neglect to make available to all thoughtful men and women therein that survey of affairs which alone can guide our way. That way will either be appallingly stern, or will be gladdened by good fortune, according as the spirit of our decision sets right at defiance, or loyally recognizes it as the compass which must guide the barque of our world's life over the ocean of the future.